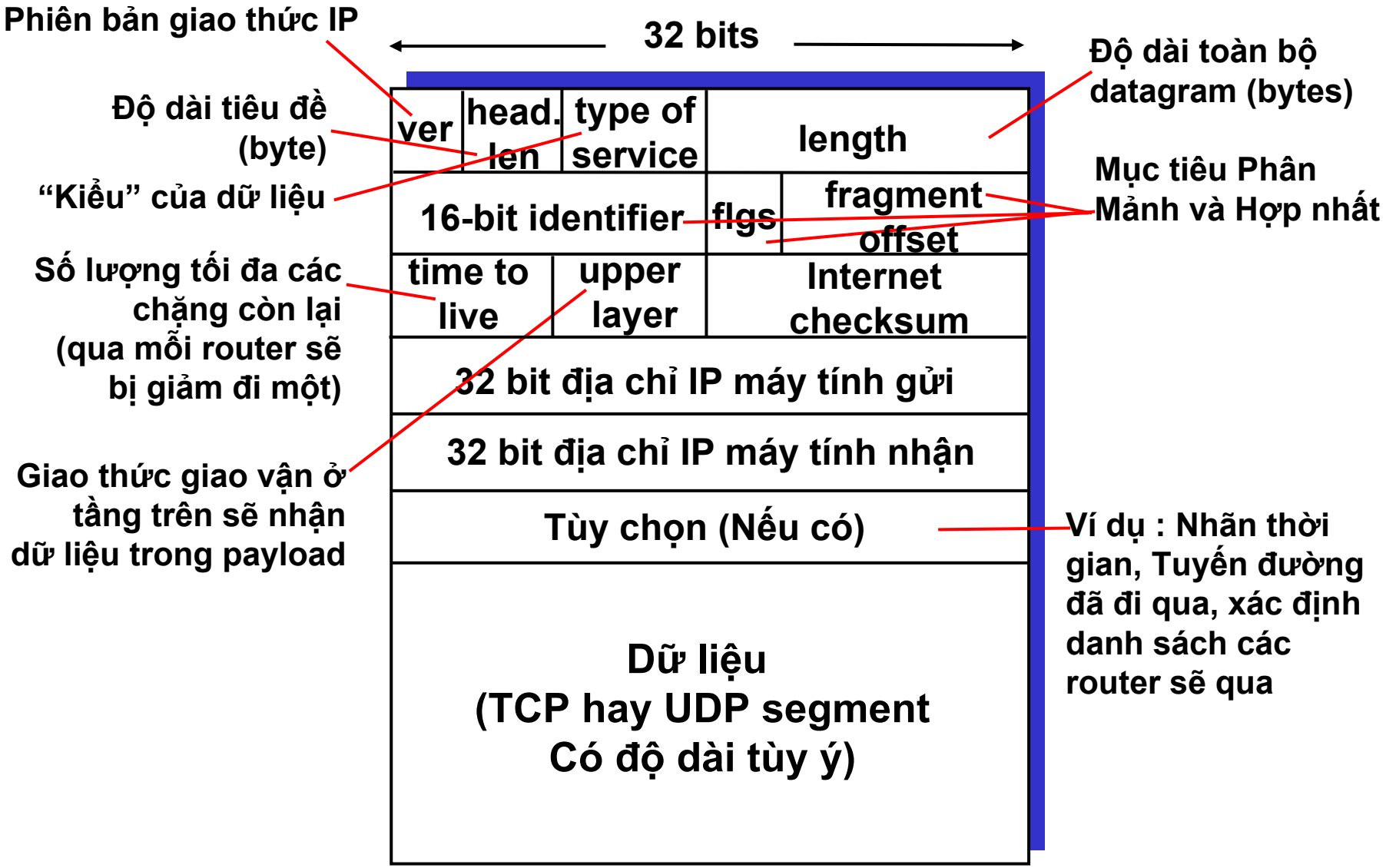
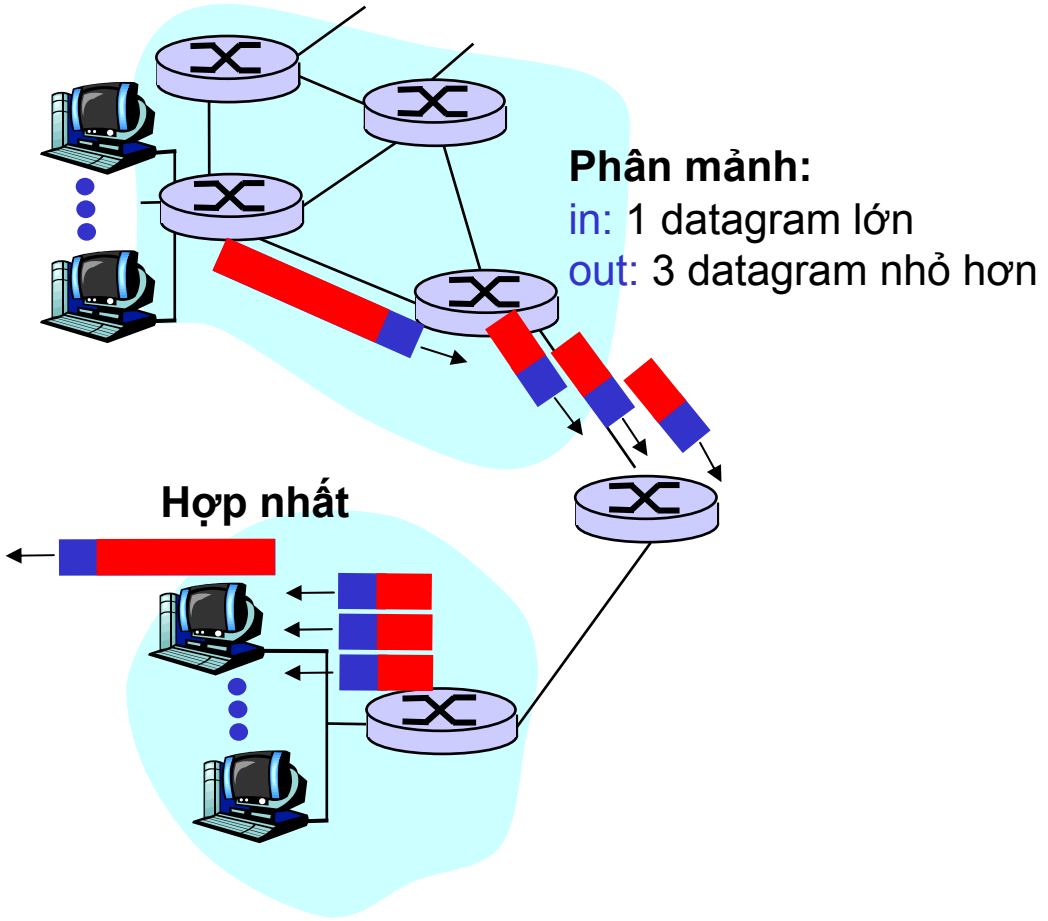
Khuôn dạng Gói tin IP

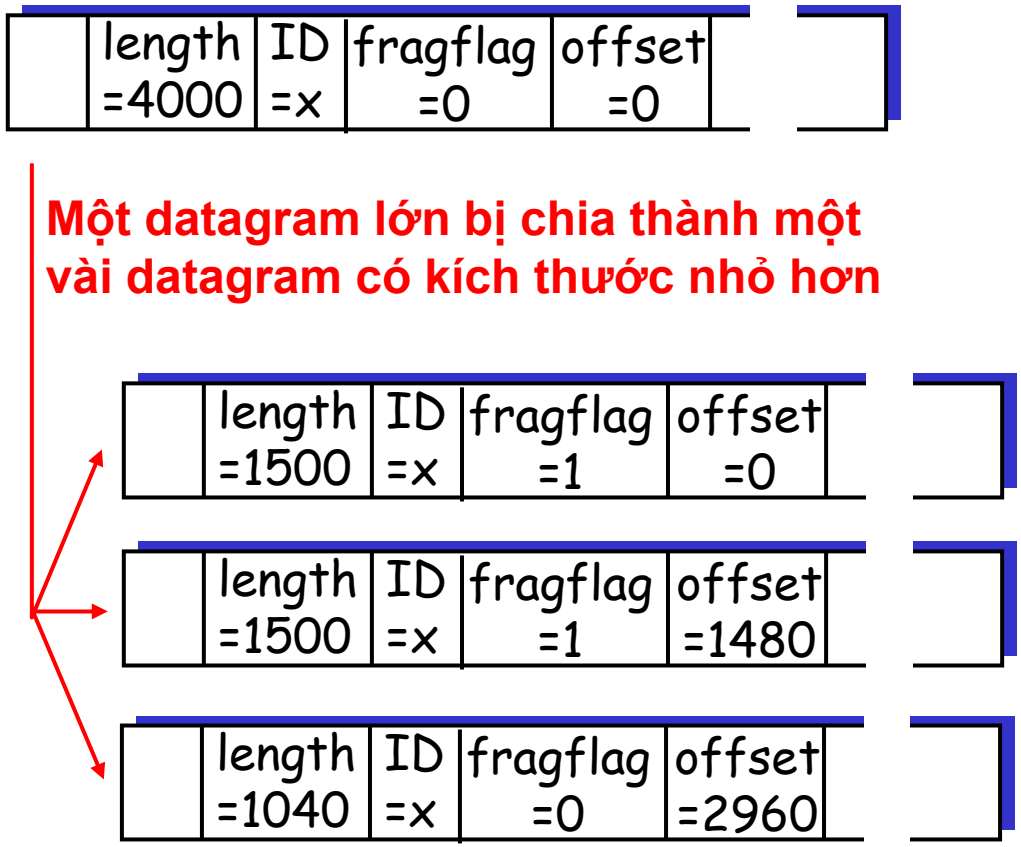


Phân mảnh và Hợp nhất Gói tin IP

- Kênh truyền có MTU (max.transfer unit) – frame lớn nhất mà tầng liên kết dữ liệu có thể gửi được.
  - Các công nghệ truyền có giá trị MTU có thể khác nhau
- IP datagram "lớn" có thể bị chia nhỏ trong mạng (Phân mảnh)
  - Một datagram bị chia thành nhiều datagram
  - Hợp nhất được thực hiện tại đích
- Các bit trong trường tiêu đề của gói tin IP được sử dụng để xác định và sắp xếp đúng thứ tự các "mảnh"



Ví dụ về Phân mảnh và Hợp nhất



ICMP: Internet Control Message Protocol

- Được các Máy tính, Router, Gateway sử dụng để trao đổi các thông tin về tầng Mạng
  - Báo lỗi: unreachable host, network, port, protocol
  - Hiển thị request/reply (Lệnh ping)
- Trong hệ thống "nằm trên" IP:
  - Thông điệp ICMP được đặt trong IP datagram
- **Thông điệp ICMP** : Kiểu (Type), Mã (code) cùng với 8 byte đầu tiên của IP datagram gây lỗi

Type	Code	description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

Chuyển tiếp Dữ liệu: Các bước

- Nếu không có lỗi, tìm kiếm địa chỉ đích của gói tin trong bảng Chuyển tiếp:
  - Nếu nút đích nằm trên mạng mà router có kết nối trực tiếp: công việc tiếp theo của tầng Liên kết dữ liệu
  - Ngược lại,
    - Tìm kiếm: xác định *router chặng kế tiếp* và giao diện ra tương ứng
    - Nếu cần thiết, phân mảnh gói tin
    - Chuyển tiếp gói tin ra giao diện của cổng ra tương ứng (là router "hàng xóm")

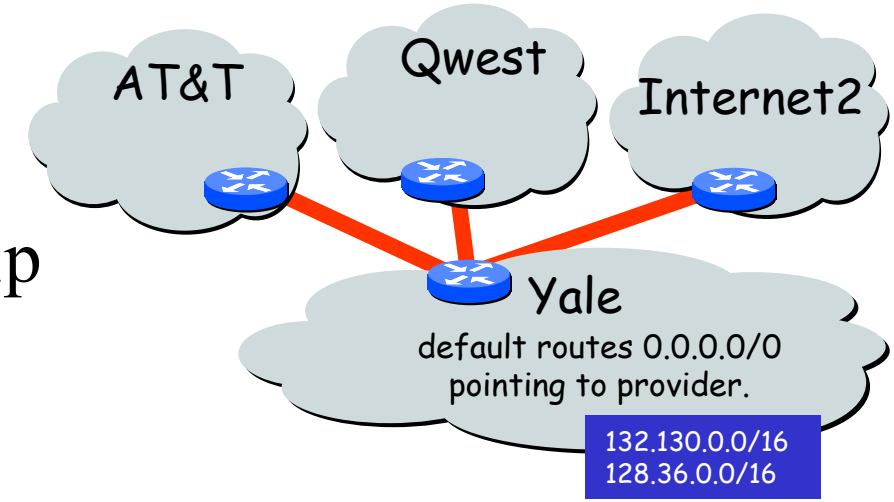
Thử %netstat –rn đê xem Bảng chuyển tiếp

Định tuyến trên Internet

- Mạng toàn cầu Internet bao gồm các Miền tự trị (Autonomous Systems - AS) kết nối với nhau:
  - Mỗi AS có một định danh riêng (ASN – AS Number)

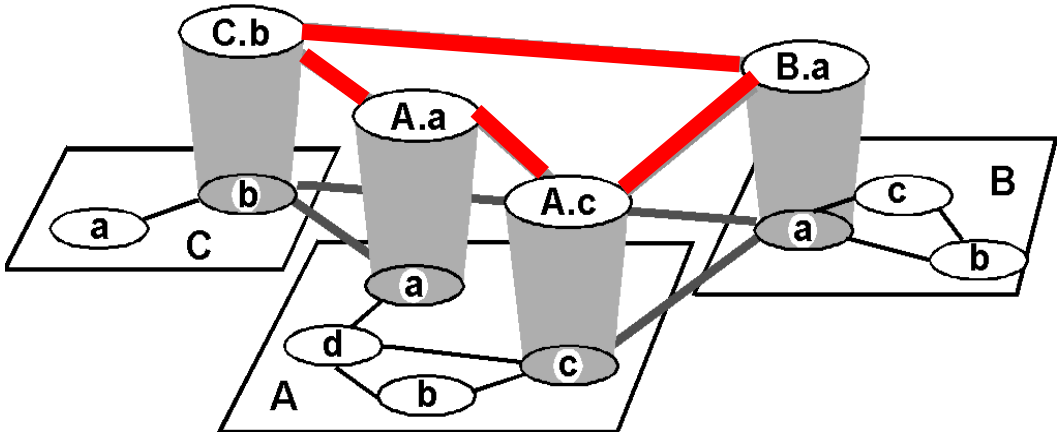
- Định tuyến hai mức:
  - **Intra-AS (Nội miền):** Người quản trị chịu trách nhiệm lựa chọn
  - **Inter-AS (Liên miền):** Chuẩn thống nhất chung

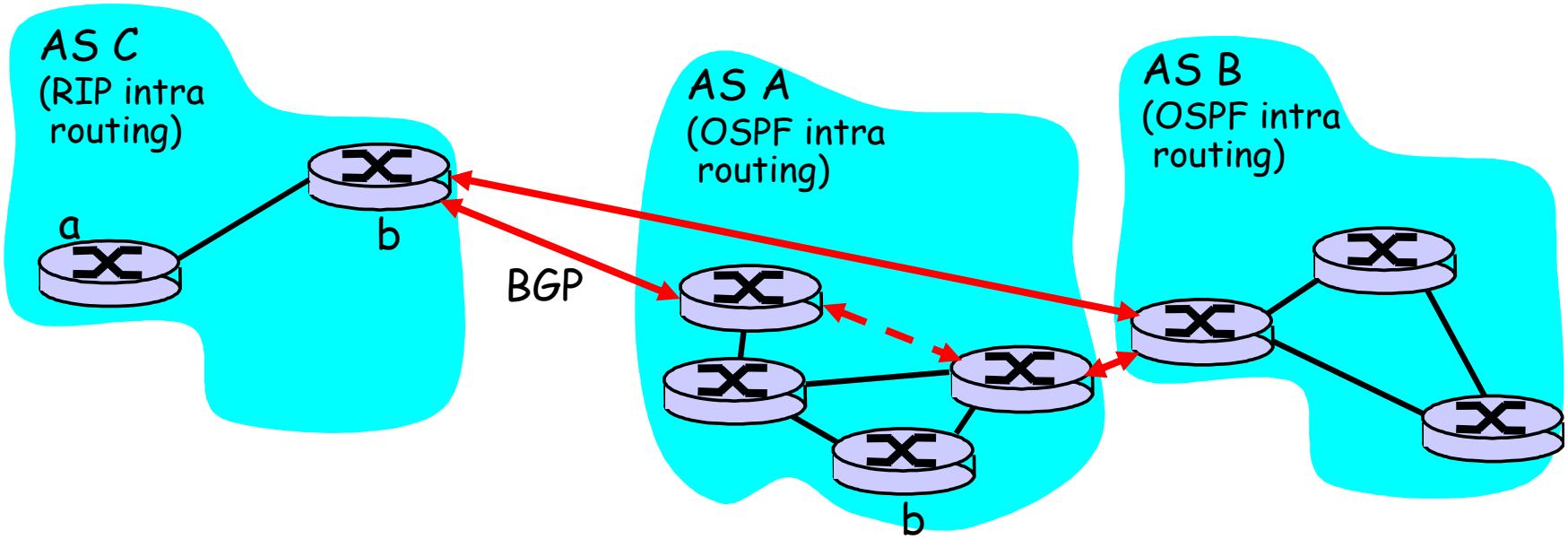
Các kiểu AS khác nhau



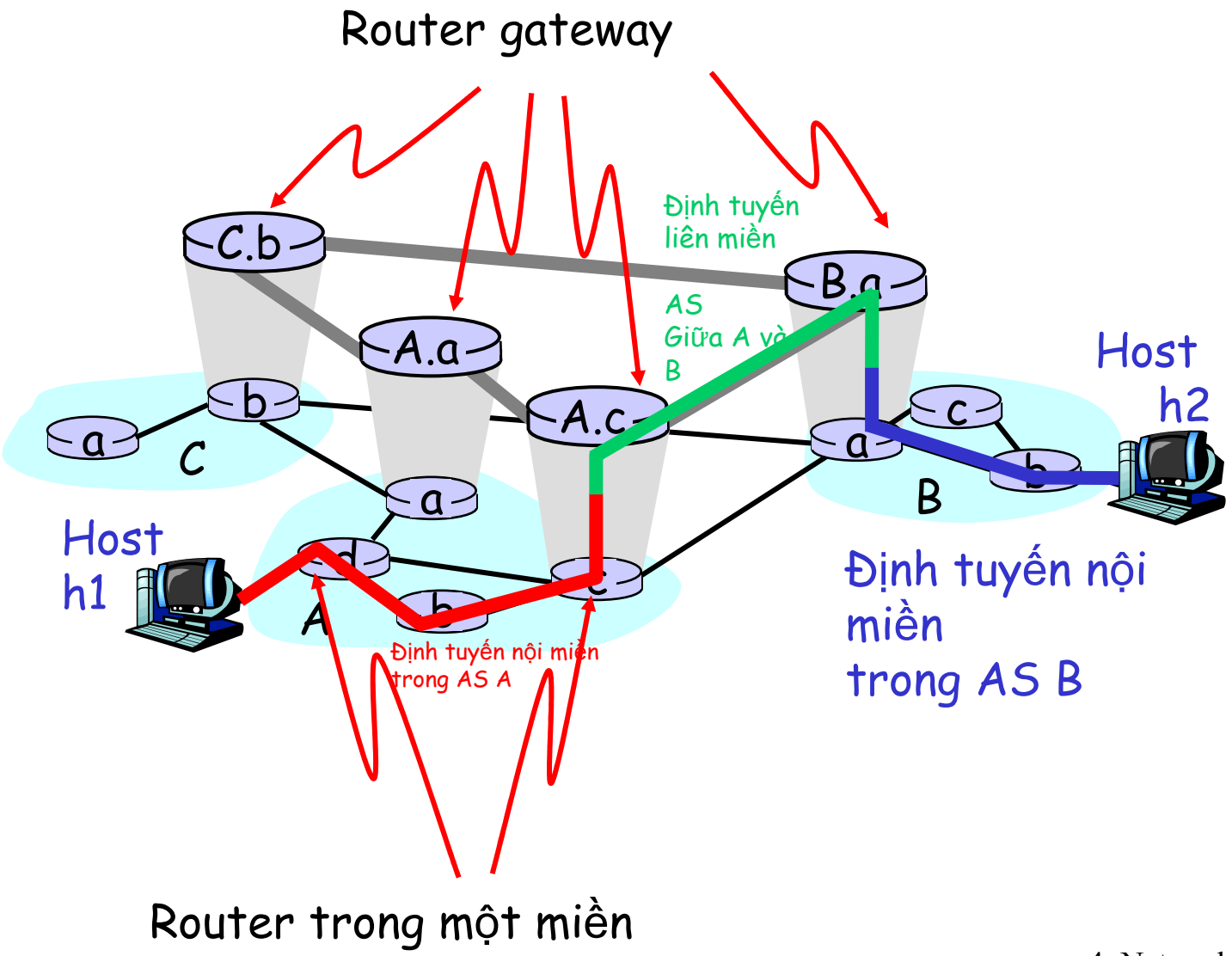
- Transit AS: Các nhà cung cấp
- Non-Transit (stub) AS
  - Phạm vi nhỏ (công ty)
- multihomed AS: Công ty lớn (Không chuyển tiếp)

Ví dụ Định tuyến trên Internet





Định tuyến Liên miền và Nội miền



Định tuyến Nội miền (Intra-AS)

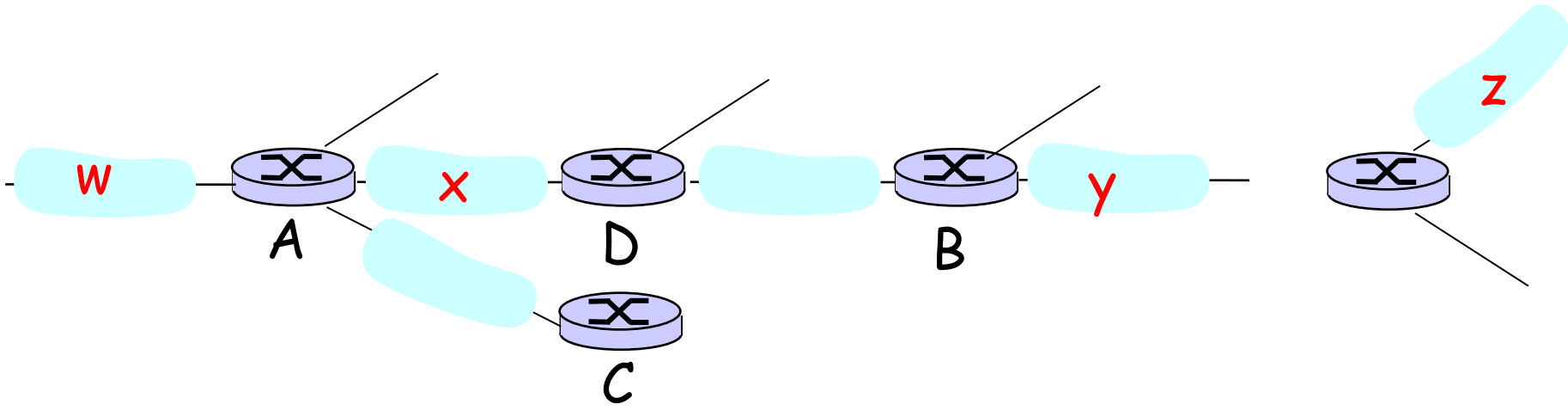
- Còn gọi là Interior Gateway Protocols (IGP)
- Một số IGP phổ biến
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (độc quyền của Cisco)

RIP (Routing Information Protocol)

- Thuật toán Distance vector
- Tích hợp trong HĐH BSD-UNIX vào năm 1982
- Đo khoảng cách: Số chặng (cực đại = 15 chặng)
  - *Tại sao như vậy?*

- Distance vectors : 30s trao đổi thông tin một lần thông qua Response Message (cũng còn gọi là **quảng cáo - advertisement**)
- Mỗi quảng cáo: có thể chuyển tới 25 trạm khác

RIP (Routing Information Protocol)



Mạng Đích	Router Kế tiếp	Số lượng các chặng đến đích.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Bảng định tuyến trong D

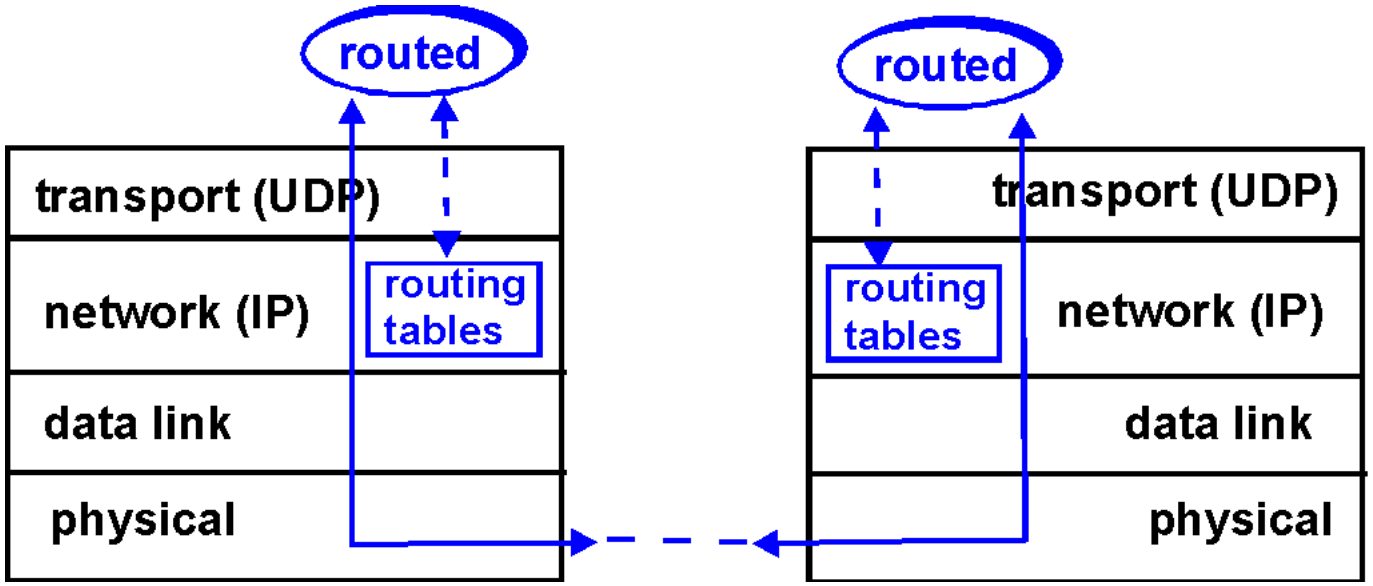
RIP: Đường truyền bị Hỏng và Khôi phục lại

Nếu không nhận được quảng cáo nào trong 180s --> Đường truyền đến hàng xóm coi như bị cắt đứt

- Tuyến đường đến hàng xóm coi như không hợp lệ
- Gửi quảng cáo đến các hàng xóm khác
- Đến lượt mình hàng xóm cũng gửi quảng cáo mới (nếu bảng định tuyến thay đổi)
- Việc một đường truyền bị hỏng nhanh chóng được các router khác biết
- poison reverse được sử dụng để ngăn chặn lặp vô hạn

Xử lý Bảng định tuyến trong RIP

- Bảng định tuyến trong RIP được tiến trình ở **tầng ứng dụng** quản lý (route-d daemon)
- Các quảng cáo được gửi định kỳ trong gói tin UDP



Ví dụ về Bảng định tuyến trong RIP

Router: *giroflee.eurocom.fr*

Destination	Gateway	Flags	Ref	Use	Interface
127.0.0.1	127.0.0.1	UH	0	26492	lo0
192.168.2.	192.168.2.5	U	2	13	fa0
193.55.114.	193.55.114.6	U	3	58503	le0
192.168.3.	192.168.3.5	U	2	25	qaa0
224.0.0.0	193.55.114.6	U	3	0	le0
default	193.55.114.129	UG	0	143454	

- Có nối với Ba mạng LAN lớp C
- Router chỉ biết các tuyến đường nối tới các LAN đó
- Router ngầm định được sử dụng để chuyển đi chỗ khác
- Địa chỉ multicast: 224.0.0.0
- Giao diện Loopback (để debug)

OSPF (Open Shortest Path First)

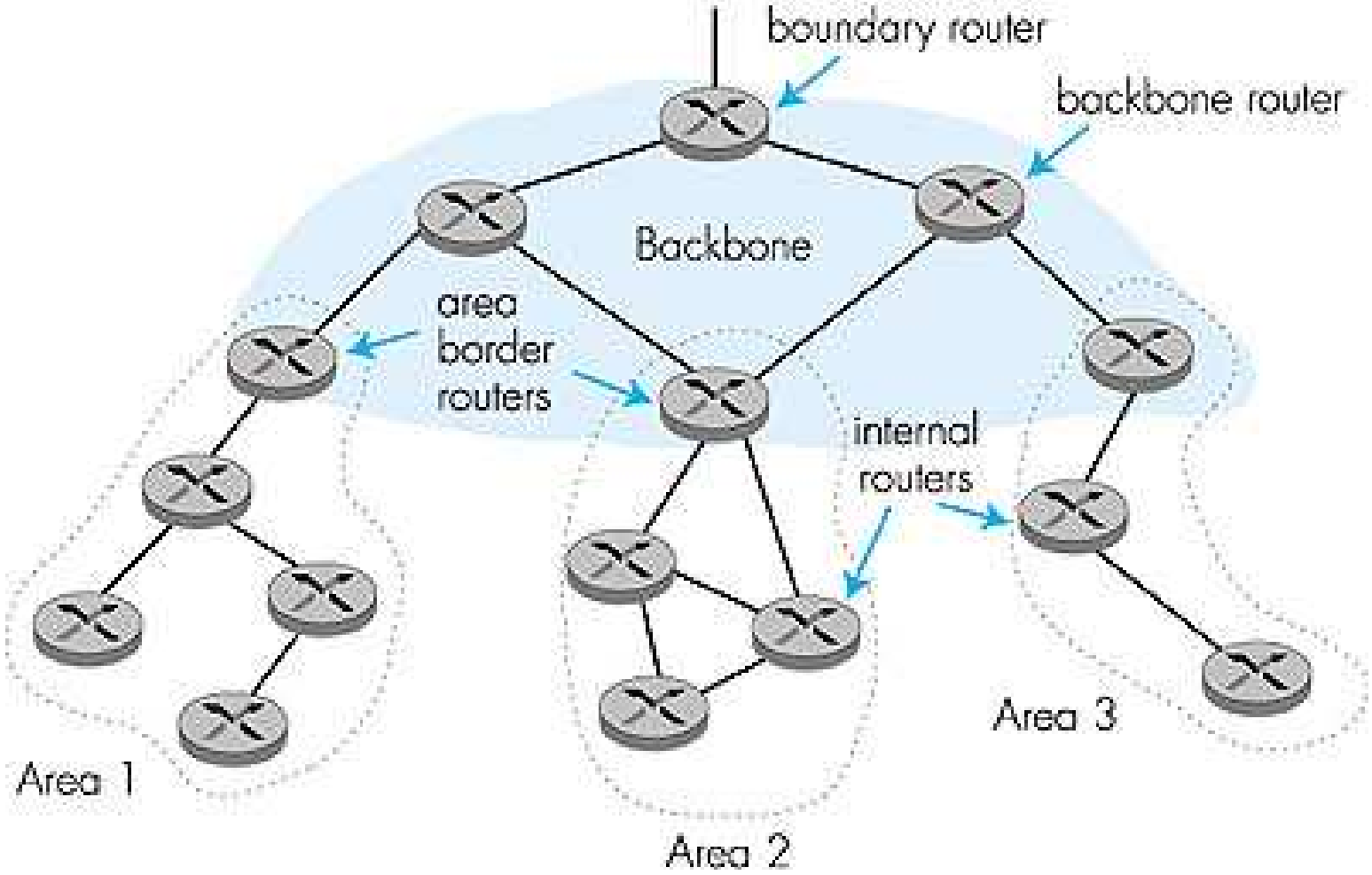
- "open": chuẩn mở
- Sử dụng thuật toán Link State
  - Gói LS được gửi
  - Nút biết về toàn bộ topo mạng
  - Tuyến đường được xác định nhờ thuật toán Dijkstra

- Thông điệp quảng cáo trong OSPF : Mỗi mục ứng với một router hàng xóm
- Các quảng cáo được gửi trên toàn bộ AS (gửi tràn ngập)

Các đặc điểm "ưu việt" của OSPF (so với RIP)

- **An ninh**: Có thể kiểm chứng các thông điệp OSPF (để ngăn ngừa phá hoại); Sử dụng kết nối TCP
- Cho phép các tuyến đường có cùng một giá (không có trong RIP)
- Trên mỗi đường truyền, có nhiều giá khác nhau cho các **TOS(Type of Service)** khác nhau (ví dụ đường truyền vệ tinh có giá "thấp" cho dịch vụ cố gắng tối đa; "cao" cho dịch vụ thời gian thực)
- Hỗ trợ gửi một đích và gửi nhiều đích (multicast):
  - Multicast OSPF (MOSPF) giống OSPF
- **Phân nhỏ** OSPF trong các miền lớn.

Phân cấp OSPF



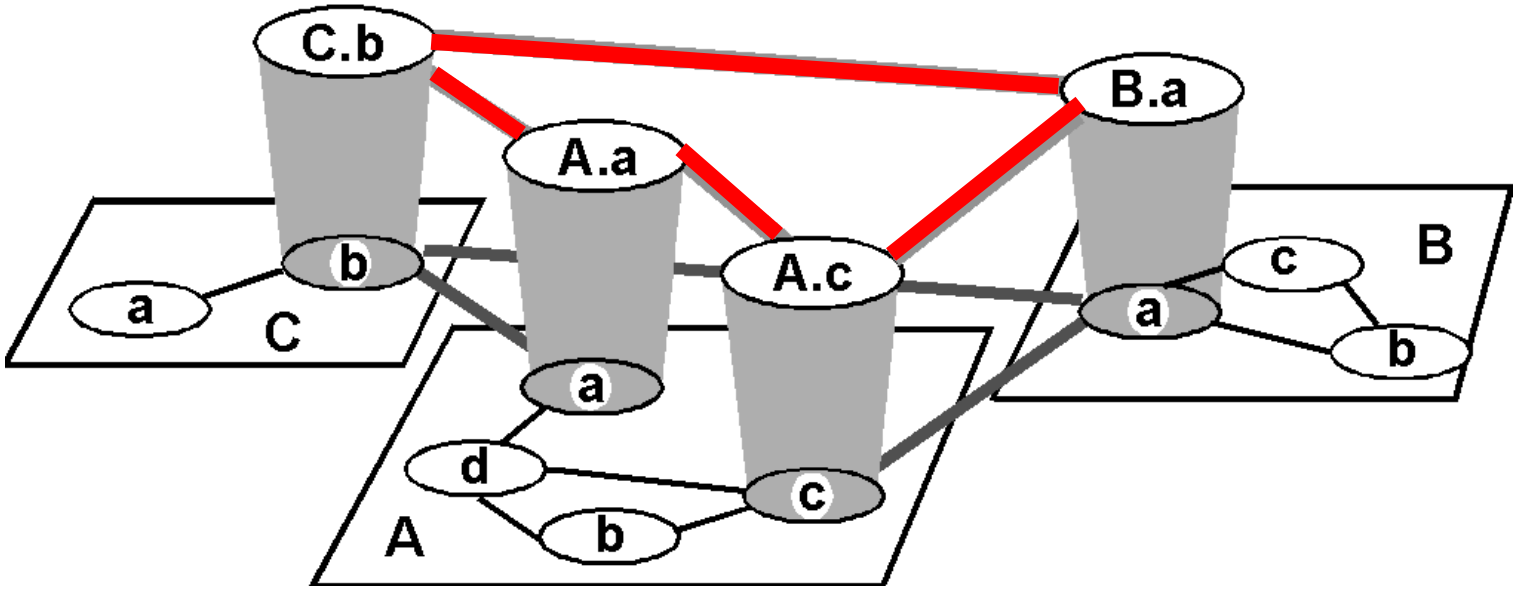
Phân cấp OSPF

- Phân cấp hai mức: cục bộ, trục chính.
  - Quảng cáo Link-state trong một vùng cục bộ
  - Mỗi nút chỉ biết topo trong một vùng; chỉ biết hướng (đường đi tốt nhất) tới các vùng khác.
- **Area border router:** "tổng hợp" khoảng cách đến các nút trong vùng, quảng cáo đến các Area Border router khác.
- **Backbone router:** chạy thuật toán định tuyến OSPF trên trục chính.
- **Boundary router:** Kết nối tới các AS khác.

IGRP (Interior Gateway Routing Protocol)

- Độc quyền của công ty CISCO; thay thế RIP (giữa 1980)
- Distance Vector, giống RIP
- Đo bằng nhiều tiêu chí khác nhau (Độ trễ, Băng thông, Độ tin cậy, Tải…)
- Sử dụng TCP để cập nhật thông tin định tuyến
- Định tuyến không bị lặp thông qua Distributed Updating Alg. (DUAL)

Định tuyến Liên miền (Inter-AS)



BGP - Định tuyến Liên miền trên Internet

- BGP (Border Gateway Protocol): *chuẩn* de facto
- Giao thức **Path Vector** :
  - Tương tự Distance Vector
  - Mỗi Border Gateway quảng bá đến các hàng xóm toàn bộ tuyến đường (là dãy các AS) tới đích
  - Ví dụ Gateway X có thể gửi đường dẫn tới đích Z:

Path (X,Z) = X,Y1,Y2,Y3,…,Z

BGP - Định tuyến Liên miền trên Internet

Giả sử: gateway X gửi tuyến đường đến gateway W

- W có thể hoặc không lựa chọn tuyến đường đi qua X
  - Chi phí, Chính sách (Không chuyển qua AS của công ty thù địch).
- Nếu W lựa chọn tuyến đường do X quảng cáo thì:

$$\text{Path (W,Z)} = w, \text{Path (X,Z)}$$

- Chú ý: X có thể kiểm soát luồng dữ liệu chuyển qua nó bằng cách kiểm soát nội dung quảng cáo gửi đến các hàng xóm:
  - Ví dụ không muốn chuyển tiếp gói tin đến Z -> không quảng cáo tuyến đường nào đến Z

BGP - Định tuyến Liên miền trên Internet

- Thông điệp BGP được đặt trong gói tin TCP.
- Thông điệp BGP :
  - OPEN: Mở kết nối TCP tới nút đối tác, kiểm chứng
  - UPDATE: Quảng cáo tuyến đường mới (hoặc xóa tuyến đường cũ)
  - KEEPALIVE Giữ tuyến đường ở trạng thái kết nối khi không gửi UPDATE; Biên nhận cho yêu cầu OPEN
  - NOTIFICATION: Báo lỗi trong thông điệp trước; cũng còn được sử dụng để đóng kết nối

Phân biệt Định tuyến Liên miền – Nội miền ?

Chính sách:

- Inter-AS: Người quản trị mong muốn kiểm soát thông tin truyền qua mạng của mình.
- Intra-AS: Một người quản trị duy nhất, do vậy không cần đến chính sách quản lý

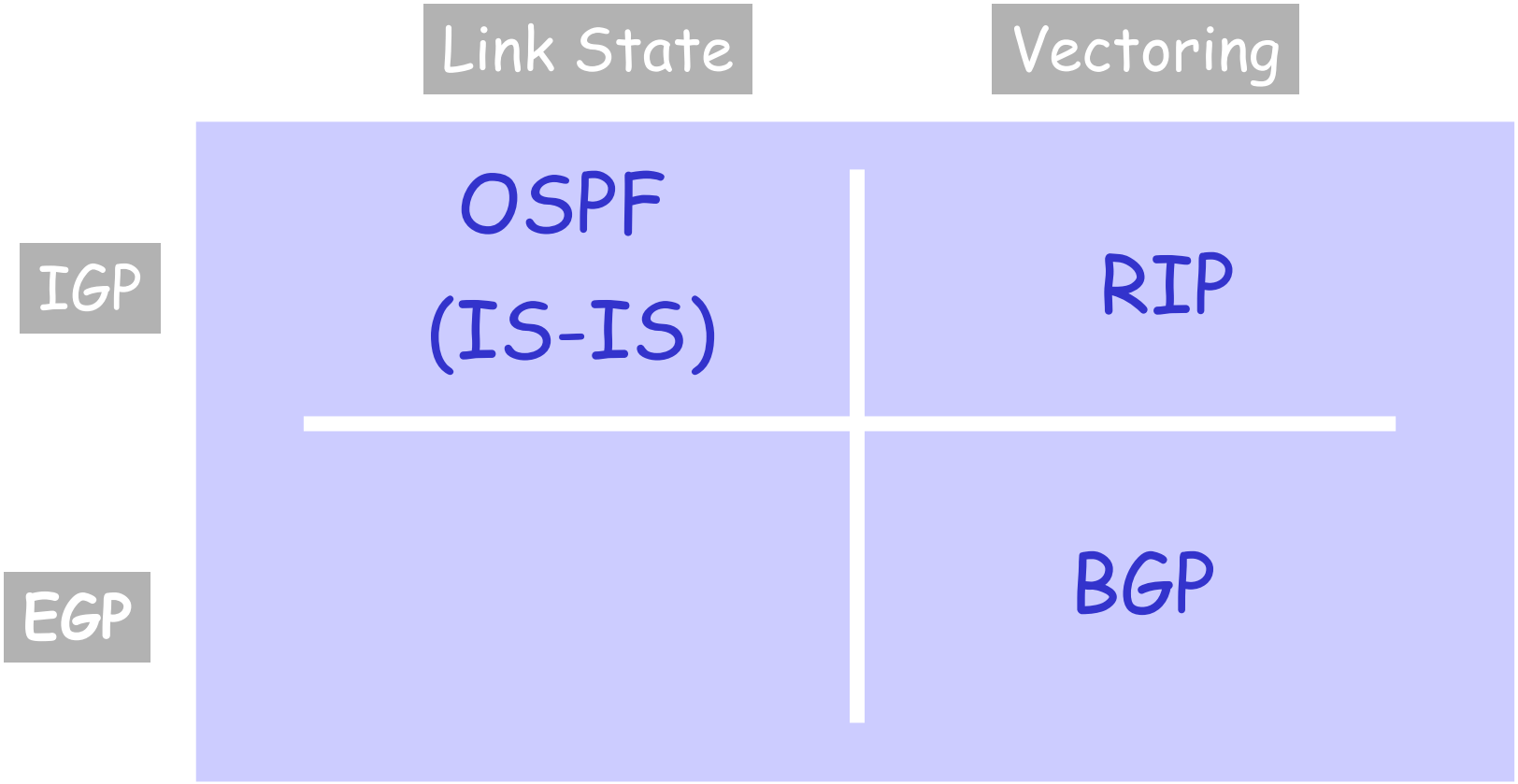
Mở rộng:

- Phân cấp giúp giảm kích thước bảng và giảm khối lượng thông tin cập nhật

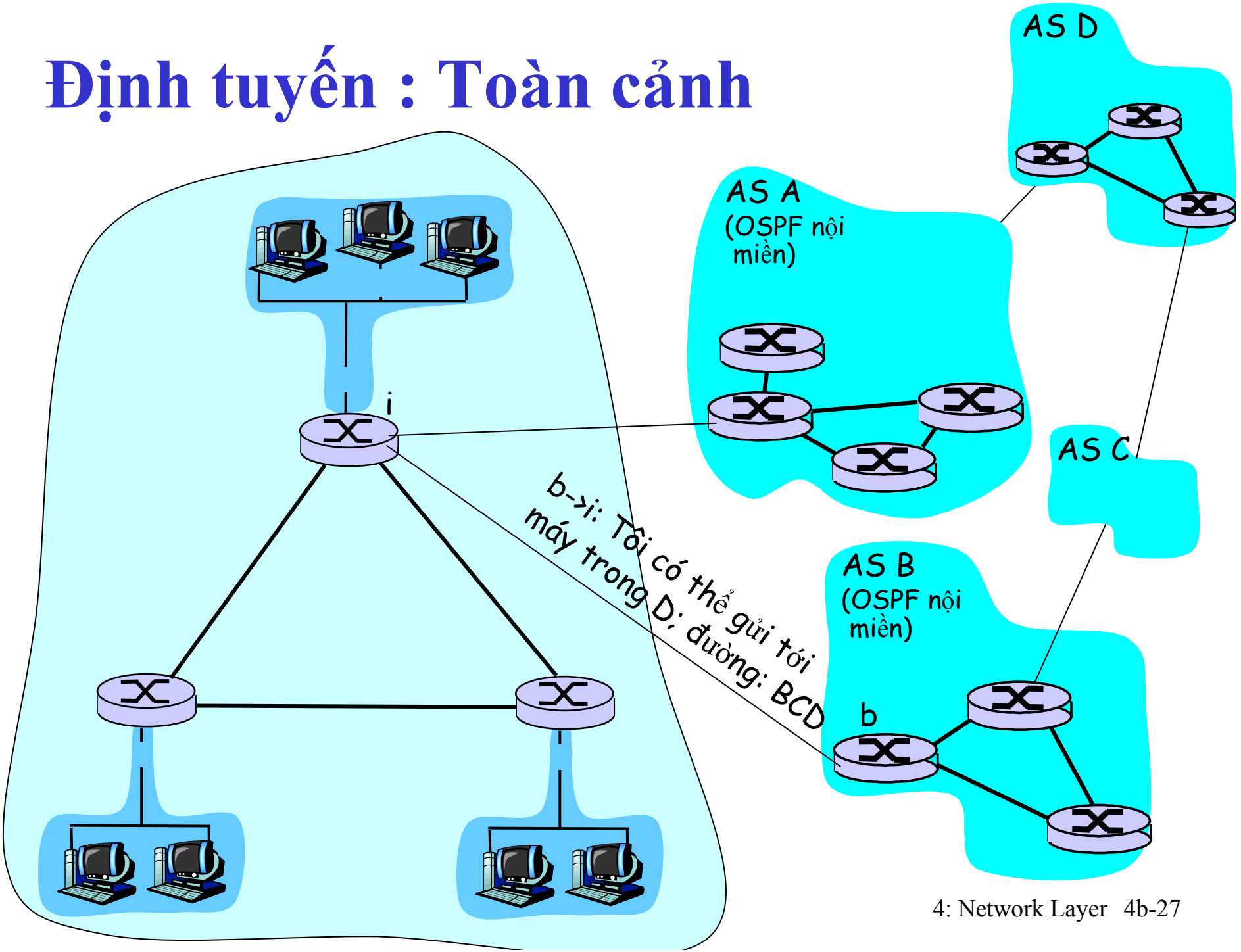
Hiệu suất:

- Intra-AS: Tập trung vào Khía cạnh Hiệu suất
- Inter-AS: Chính sách có thể được ưu tiên hơn Hiệu suất

Summary: The Gang of Four

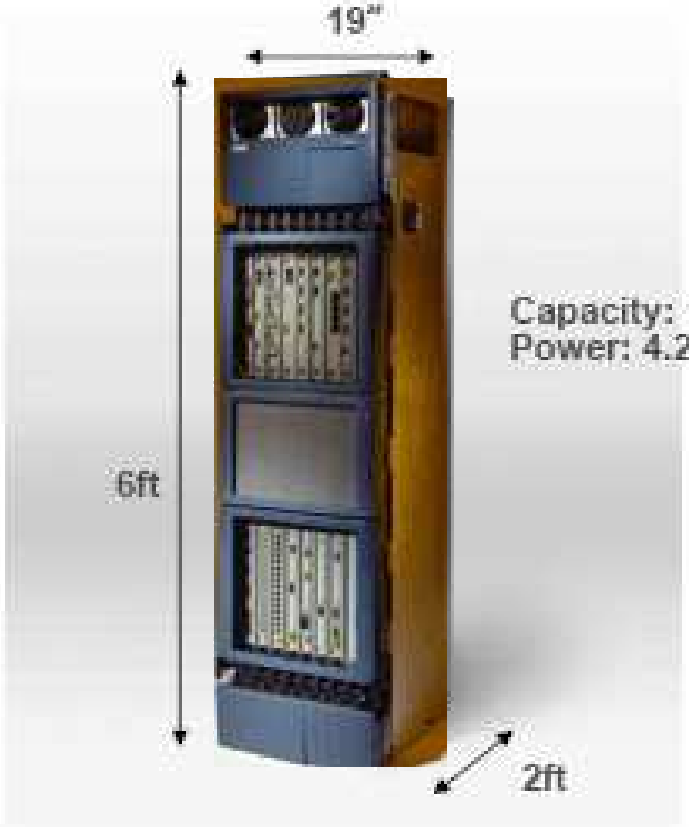


Định tuyến : Toàn cảnh



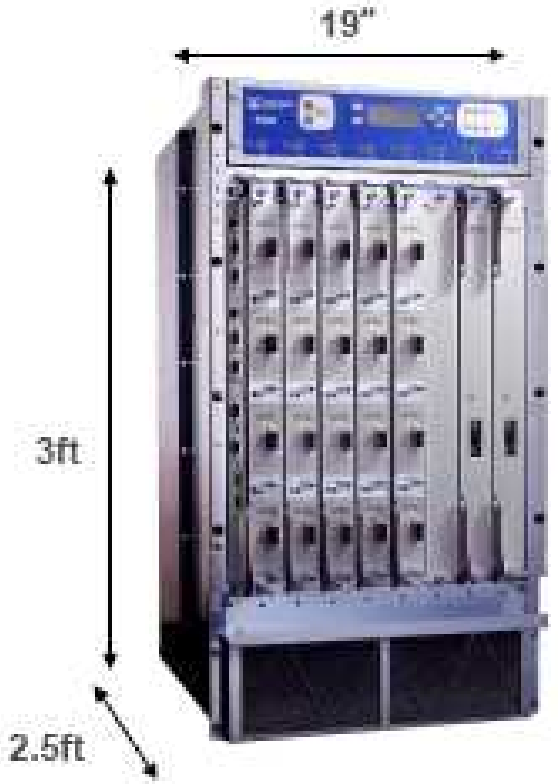
Router trông như thế nào ?

Cisco GSR 12416



Capacity: 160Gb/s
Power: 4.2kW

Juniper M160

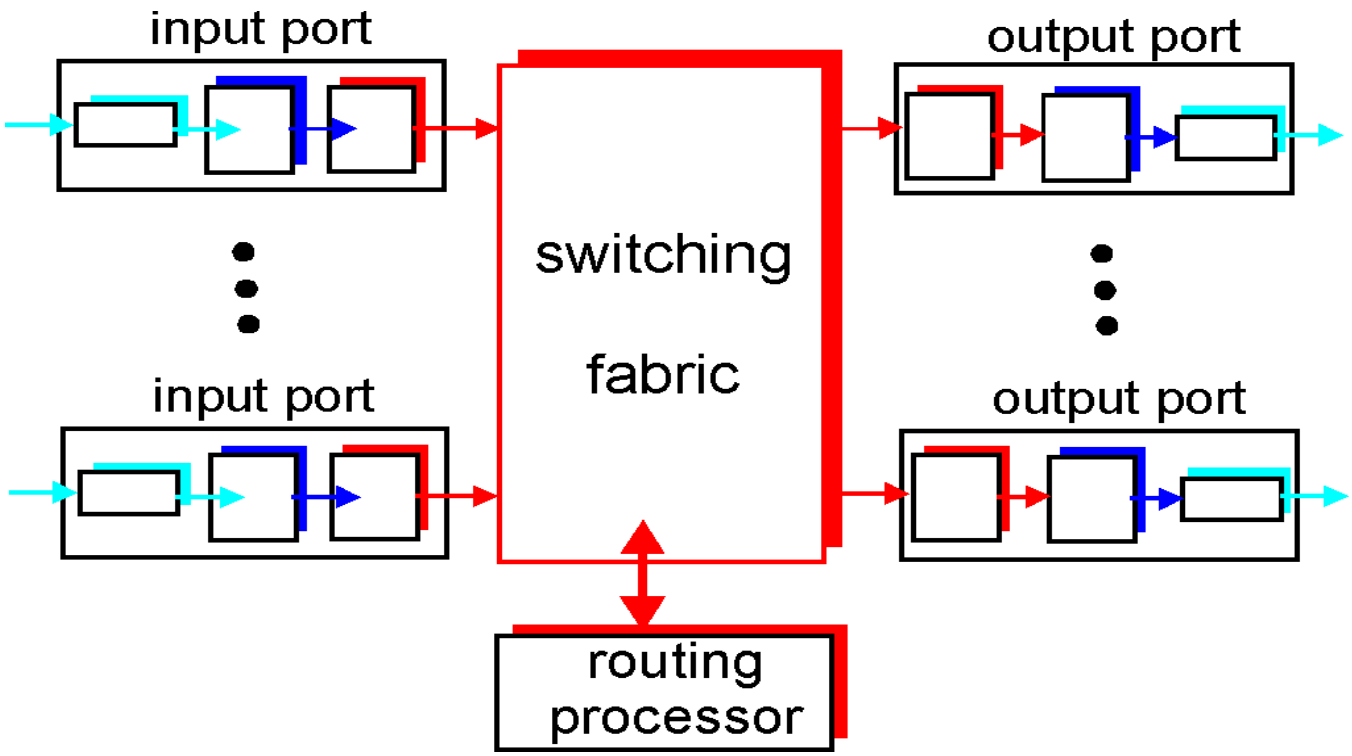


Capacity: 80Gb/s
Power: 2.6kW

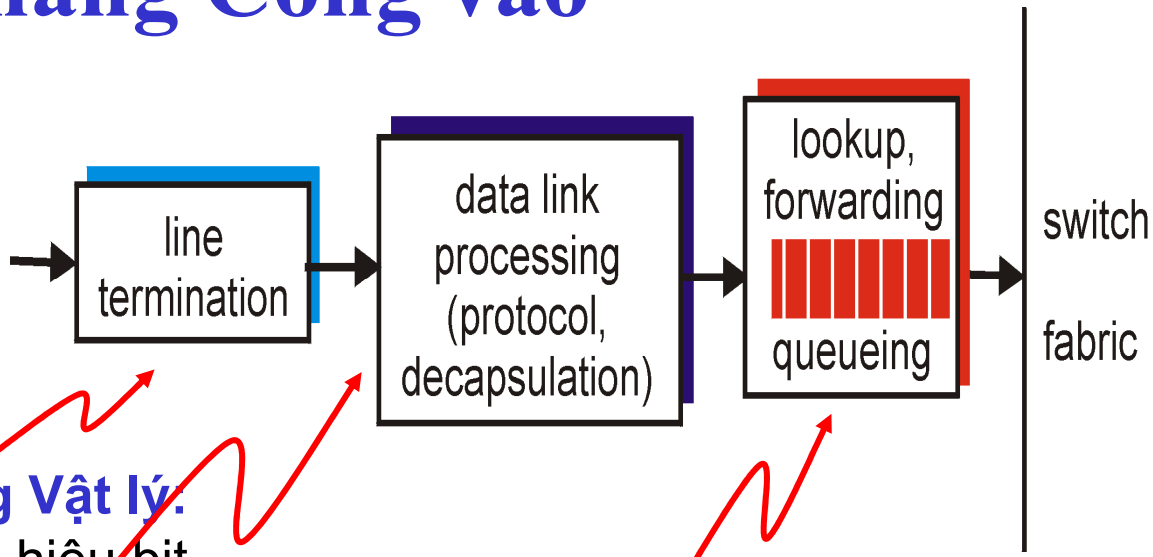
Tổng quan Kiến trúc của Router

Hai chức năng chính:

- Chạy các thuật toán, giao thức định tuyến (RIP, OSPF, BGP)
- *Chuyển* datagram từ cổng vào đến cổng ra thích hợp



Chức năng Cổng vào



Tầng Vật lý:
Nhận tín hiệu bit

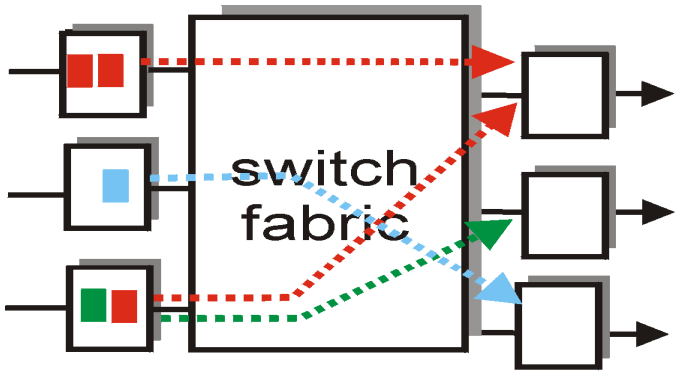
Liên kết Dữ liệu
Ví dụ Ethernet (chương 5)

Chuyển Không tập trung

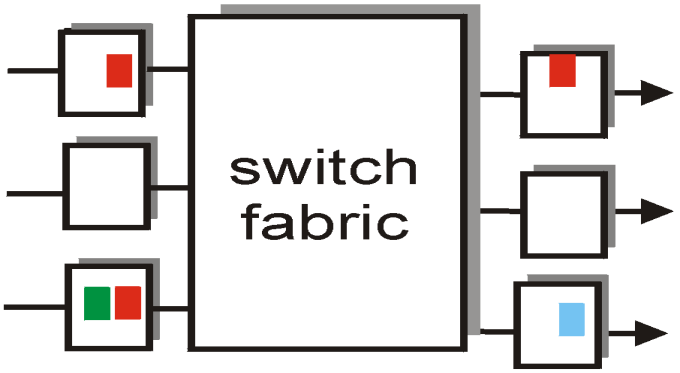
- Với một địa chỉ đích, tìm kiếm trên bảng định tuyến để xác định cổng ra phù hợp
- Mục tiêu: Thực hiện xử lý tại cổng vào ở tốc độ đến của gói tin
- Hàng đợi: Xuất hiện khi tốc độ đến của gói tin nhanh hơn khả năng chuyển đi của Kết cấu chuyển

Hàng đợi tại Cổng vào

- Kết cấu chuyển chậm hơn tổng lượng đến từ các cổng vào -> Xuất hiện Hàng đợi ở một số cổng
- Phong tỏa Đầu Hàng đợi (Head-of-the-Line HOL) : datagram ở đầu hàng đợi bị phong tỏa ngăn chặn các datagram sau
- *Độ trễ hay Mất do xếp hàng Vì tràn bộ đệm tại Cổng vào!*

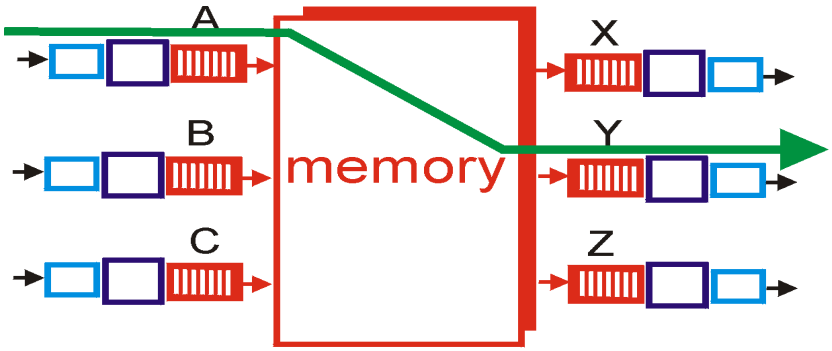


output port contention
at time t - only one red
packet can be transferred

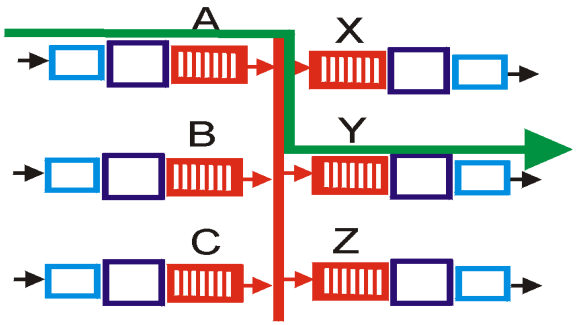


green packet
experiences HOL blocking

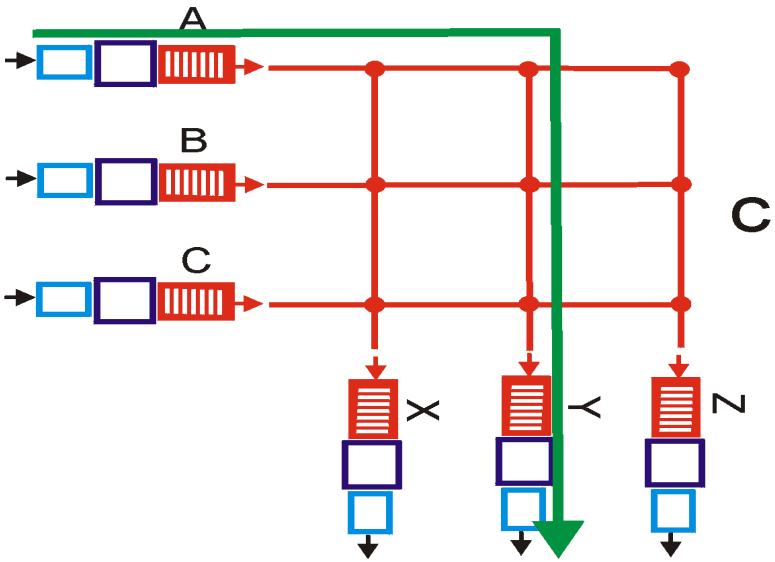
Ba kiểu Kết cấu chuyển



memory



bus

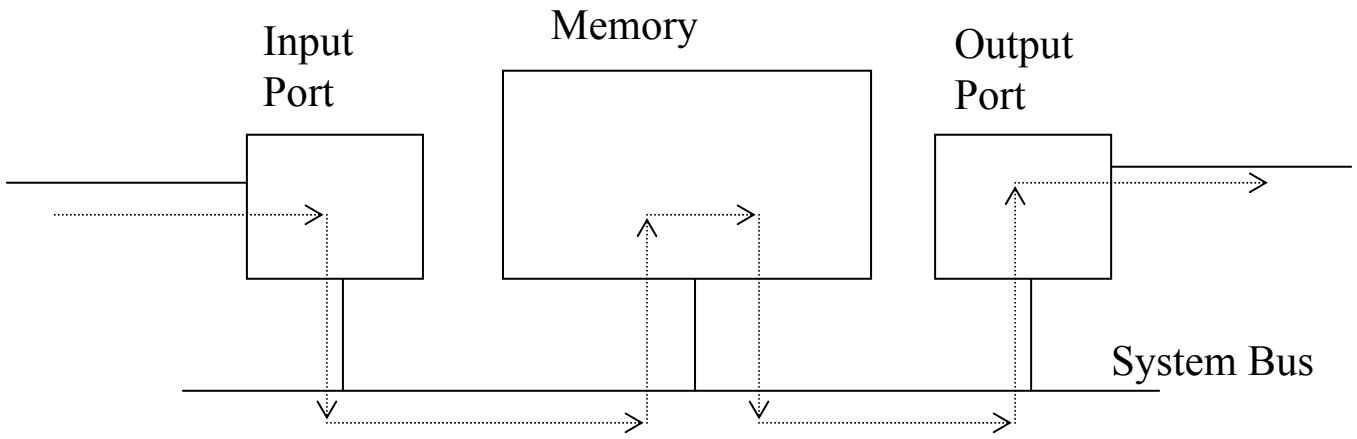


crossbar

Chuyển qua Bộ nhớ

Các Router thế hệ đầu tiên:

- CPU (duy nhất) thực hiện sao chép các packet
- Tốc độ bị giới hạn bởi băng thông bộ nhớ (mỗi datagram cần 2 lần sử dụng bus)



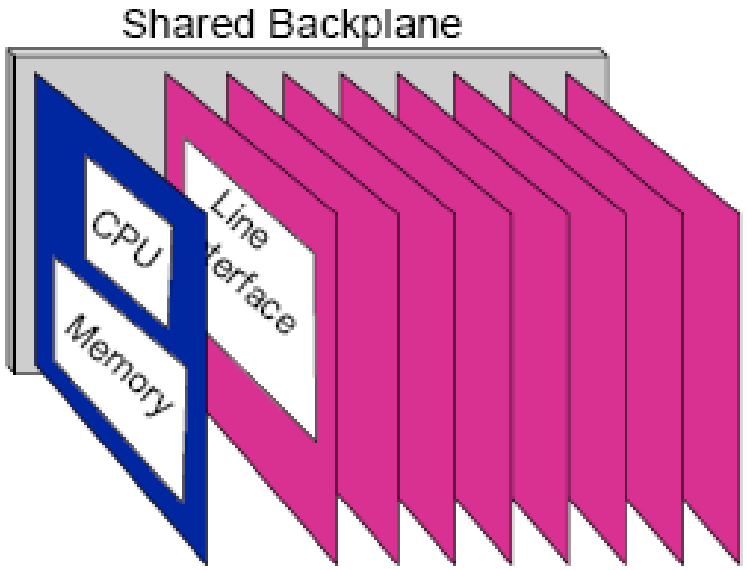
Router hiện đại:

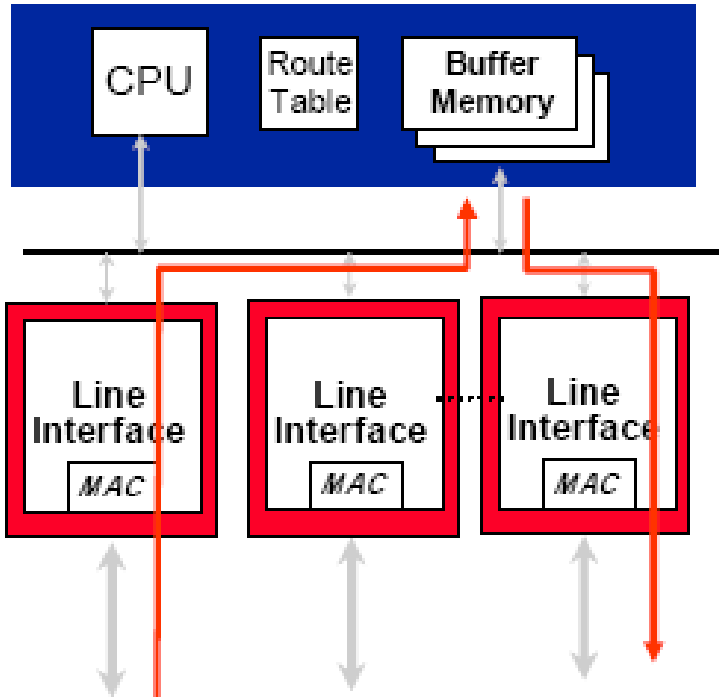
- CPU tại cổng vào thực hiện tìm kiếm và chuyển vào bộ nhớ
- Cisco Catalyst 8500

Chuyển qua Bộ nhớ

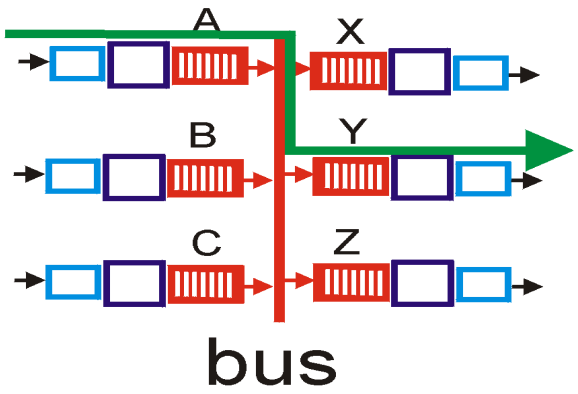
Thế hệ router đầu tiên: CPU duy nhất của máy tính di chuyển các packet

- Nút cổ chai: Bộ nhớ dùng chung; datagram chuyển qua bus 2 lần





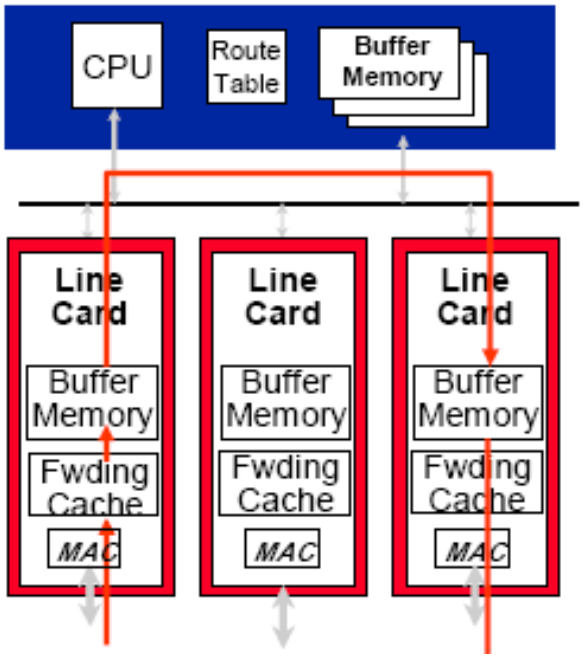
Chuyển qua Bus



- datagram chuyển từ Bộ nhớ Cổng vào đến Bộ nhớ Cổng ra qua bus dùng chung
- Tranh chấp bus: Tốc độ chuyển mạch bị giới hạn bởi băng thông của bus
- 1 Gbps bus, Cisco 1900: Tốc độ đủ cao cho các router của công ty

Chuyển qua Bus

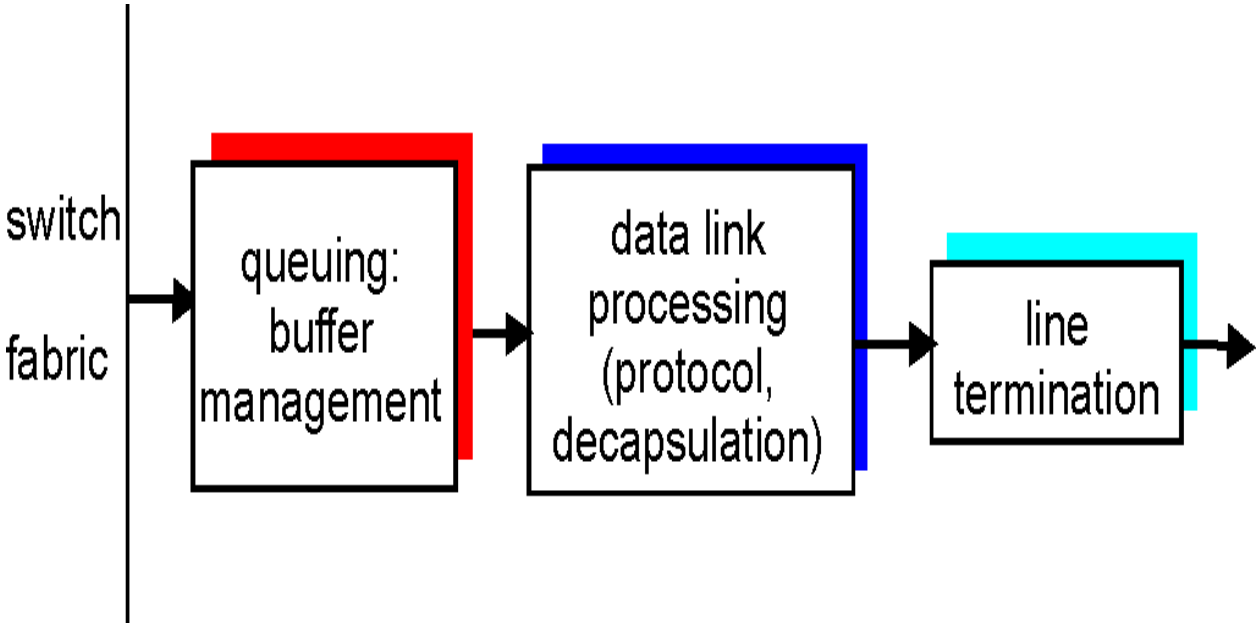
- Datagram từ Bộ nhớ cổng vào tới Bộ nhớ cổng ra qua bus dùng chung
- Nút cổ chai: tranh chấp bus
  - < 5Gbps, ví dụ 1 Gbps bus, Cisco 1900: tốc độ truy cập vừa đủ cho router của công ty (không phải router trên các trục chính)



Chuyển qua Mạng liên hợp

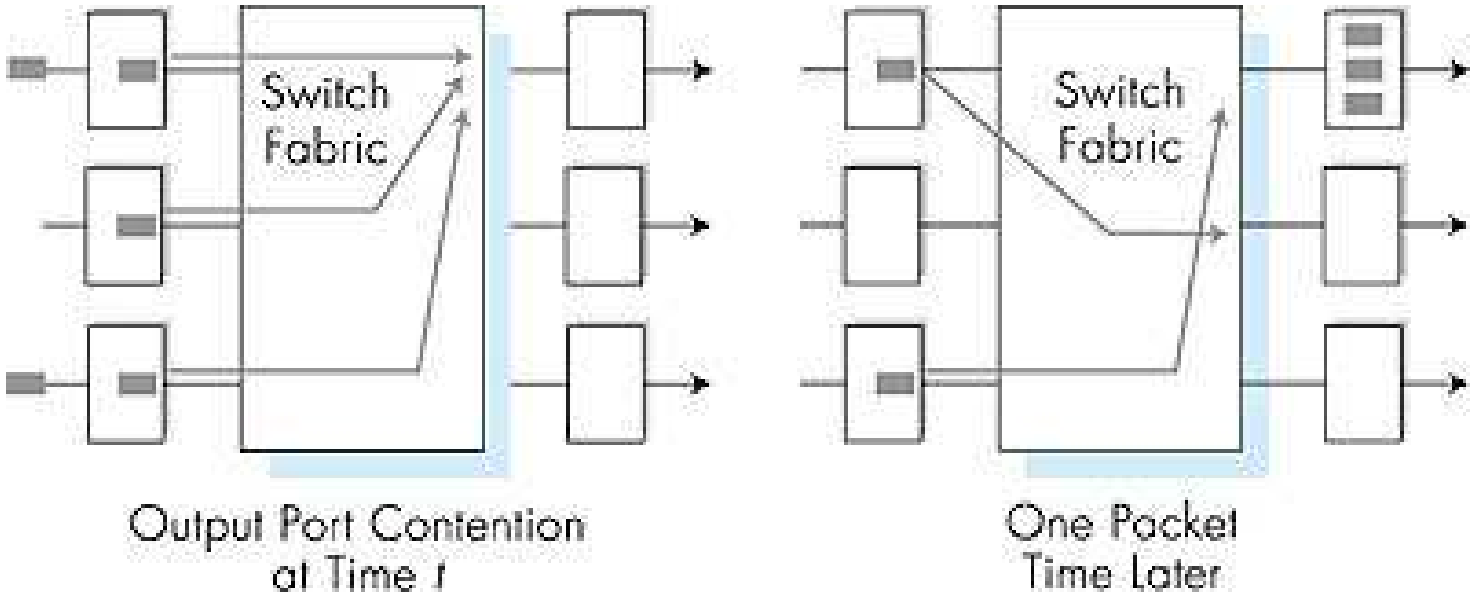
- Khắc phục hạn chế Băng thông của Bus
- Mạng Banyan, Khởi đầu để kết nối các CPU trong hệ thống có nhiều CPU
- Thiết kế cao cấp: chia datagram thành các "tế bào" có kích thước cố định và chuyển các "tế bào" qua kết cấu chuyển.
- Cisco 12000: Tốc độ Gbps qua mạng kết cấu chuyển

Cổng Ra



- ***Bộ đệm*** : Cần thiết trong trường hợp khi gói tin đến từ Kết cấu chuyển nhanh hơn Tốc độ gửi của Kênh truyền
- ***Chiến lược điều phối*** lựa chọn datagram để chuyển trong các datagram nằm ở Bộ đệm

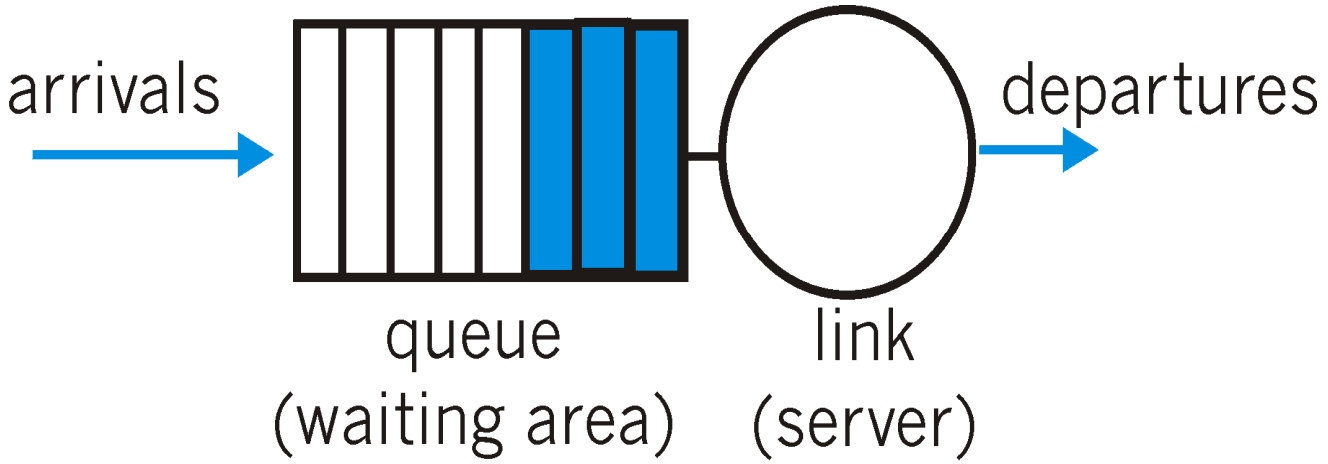
Hàng đợi tại Cổng ra



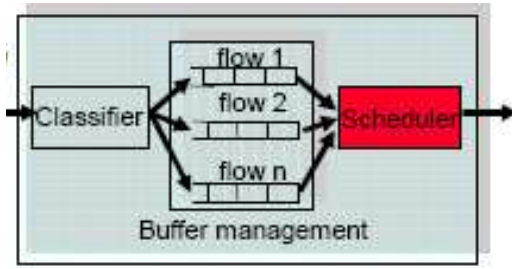
- Tốc độ gửi chậm hơn tốc độ kết cấu chuyển
- *Xếp hàng (Trễ) và Mất dữ liệu do **tràn Bộ đệm** ở Cổng ra!*

Cơ chế Điều phối

- Điều phối: Chọn packet kế tiếp để chuyển đi trên đường truyền
- Điều phối FIFO (first in first out) : Gửi theo thứ tự đến Hàng đợi

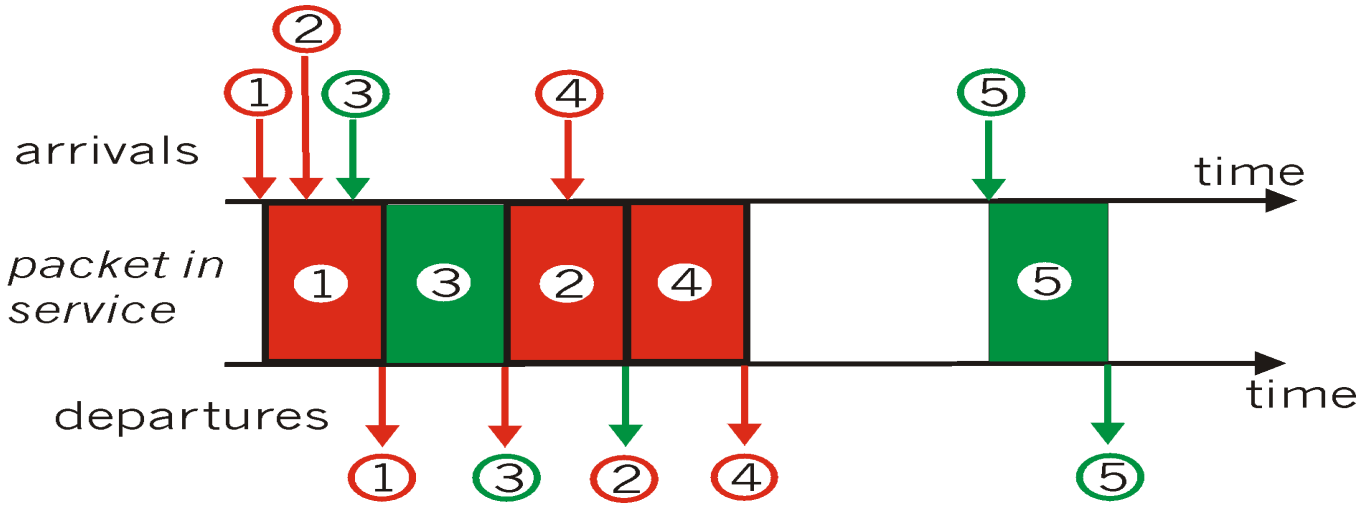


Các chính sách điều phối khác



Điều phối xoay vòng (Round Robin):

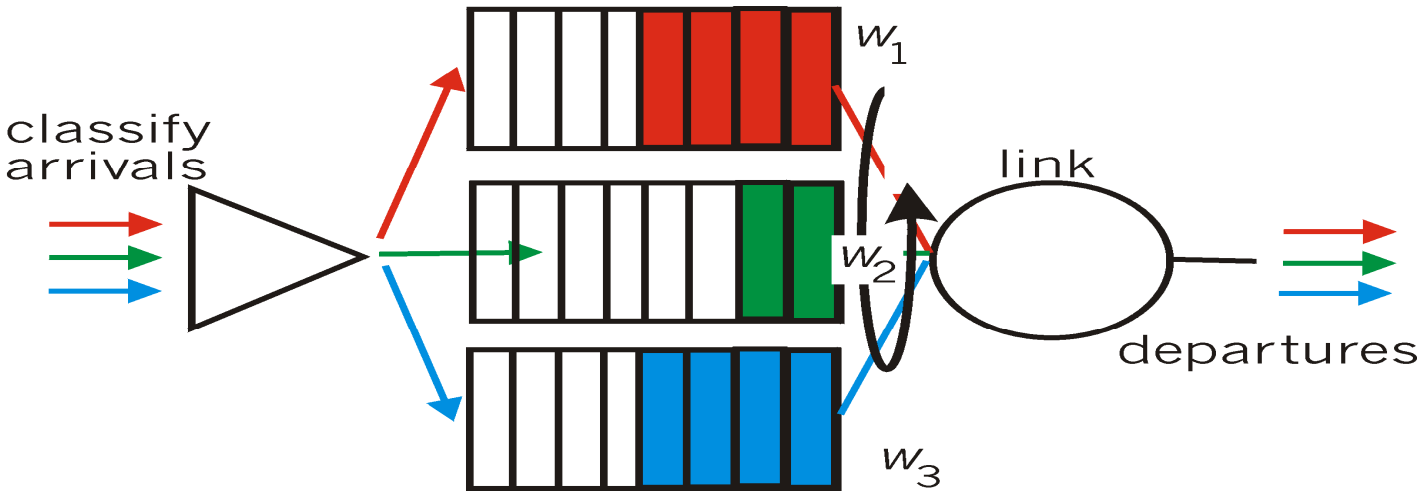
- Chia ra nhiều lớp
- Lần lượt và xoay vòng phục vụ từng hàng đợi



Các chính sách điều phối khác

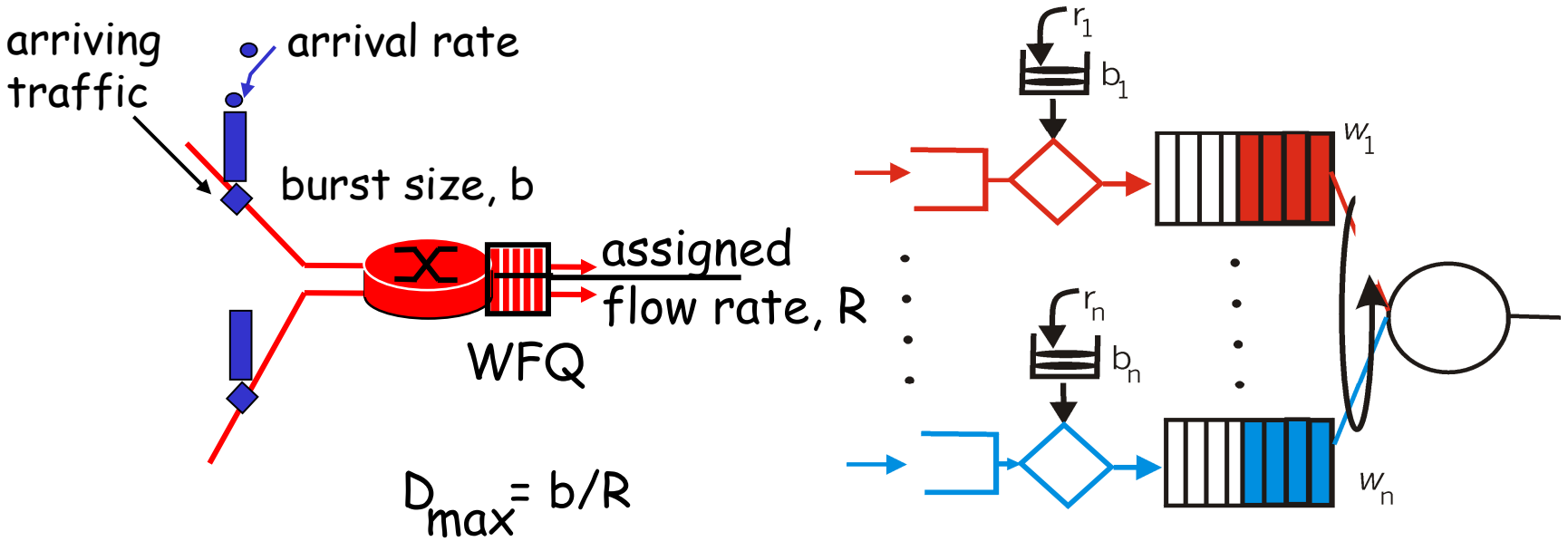
Theo trọng số (Weighted Fair Queuing):

- Tổng quát hóa Round Robin
- Mỗi lớp có trọng số phục vụ riêng



Đảm bảo giới hạn Độ trễ

- WFQ đảm bảo cận trên của độ trễ tức là vấn đề *Đảm bảo Chất lượng Dịch vụ (QoS guarantee)*!



IPv6

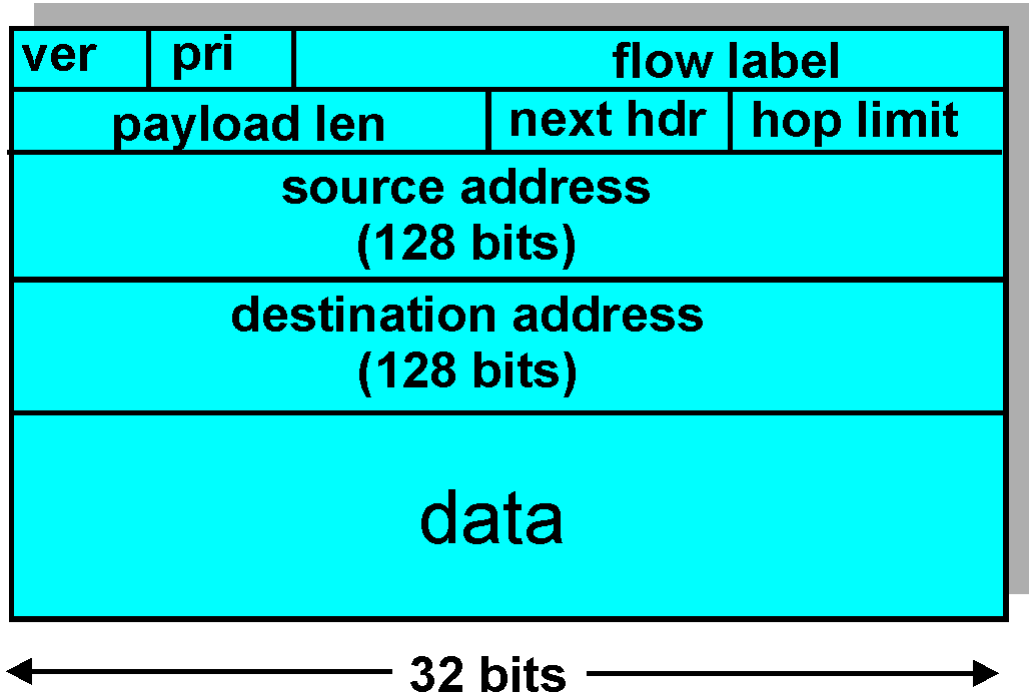
- **Động lực đầu tiên**: Không gian địa chỉ 32-bit sẽ cạn kiệt vào 2008.
- Các động lực khác:
  - Khuôn dạng của tiêu đề ảnh hưởng đến tốc độ Xử lý và Chuyển tiếp gói tin
  - Thay đổi tiêu đề để đáp ứng Chất lượng Dịch vụ (QoS)
  - Địa chỉ kiểu "anycast" : tuyến đường "tốt nhất" trong một vài server nhân bản
- **Khuôn dạng gói IPv6 :**
  - Tiêu đề có độ dài cố định 40 byte
  - Không cho phép Phân mảnh

Tiêu đề của IPv6

Priority: Xác định độ ưu tiên trong luồng dữ liệu

Flow Label: xác định các datagrams trong cùng một "luồng" (khái niệm "luồng" chưa rõ ràng).

Next header: Xác định giao thức giao vận nhận dữ liệu



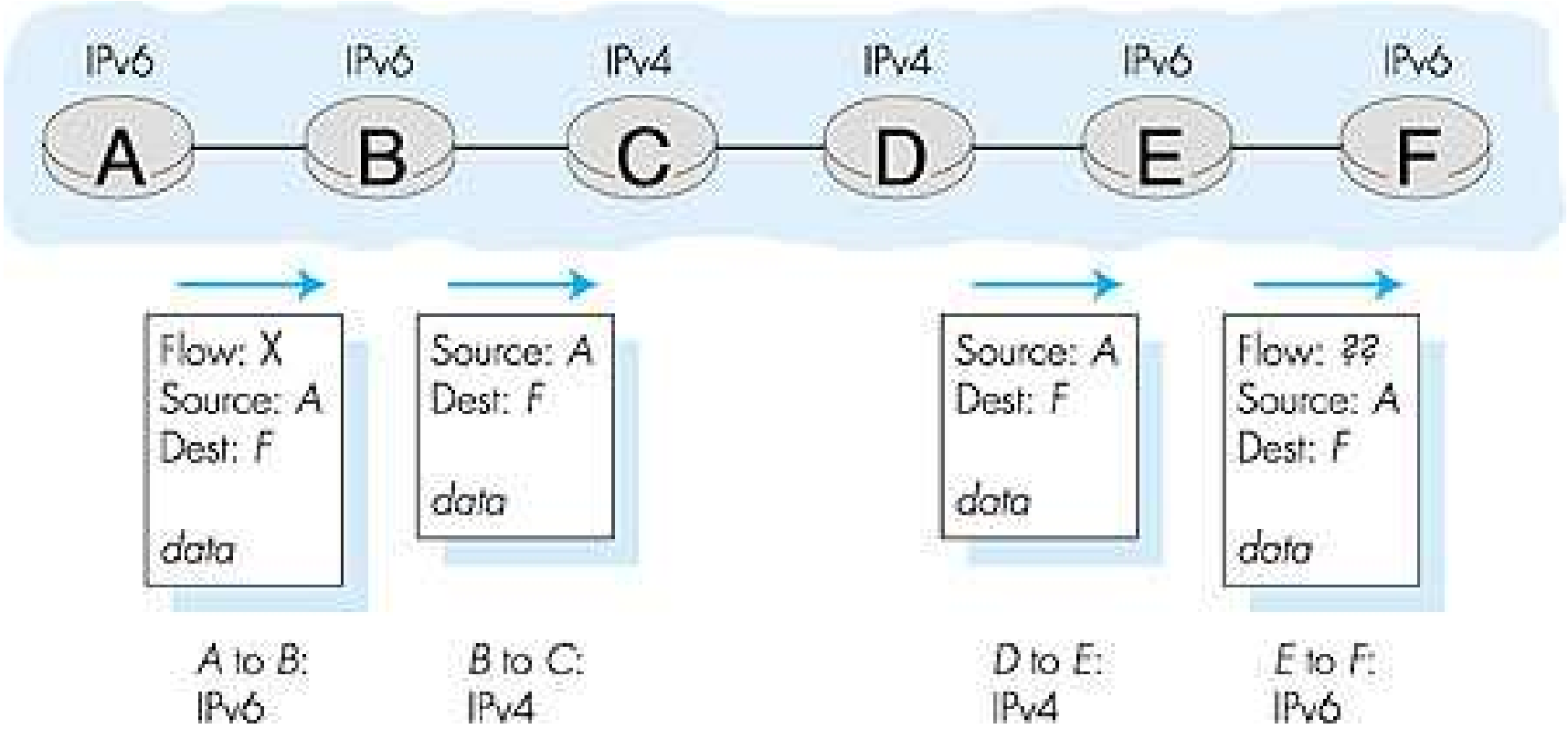
Khác biệt so với IPv4

- *Checksum*: Bị loại bỏ để tăng tốc độ xử lý gói tin tại mỗi router
- *Options:* cho phép, nhưng nằm ngoài tiêu đề, được xác định qua trường "Next Header"
- *ICMPv6:* Phiên bản mới của ICMP
  - Các kiểu thông điệp mới, ví dụ "Packet Too Big"
  - Chức năng quản lý nhóm Multicast

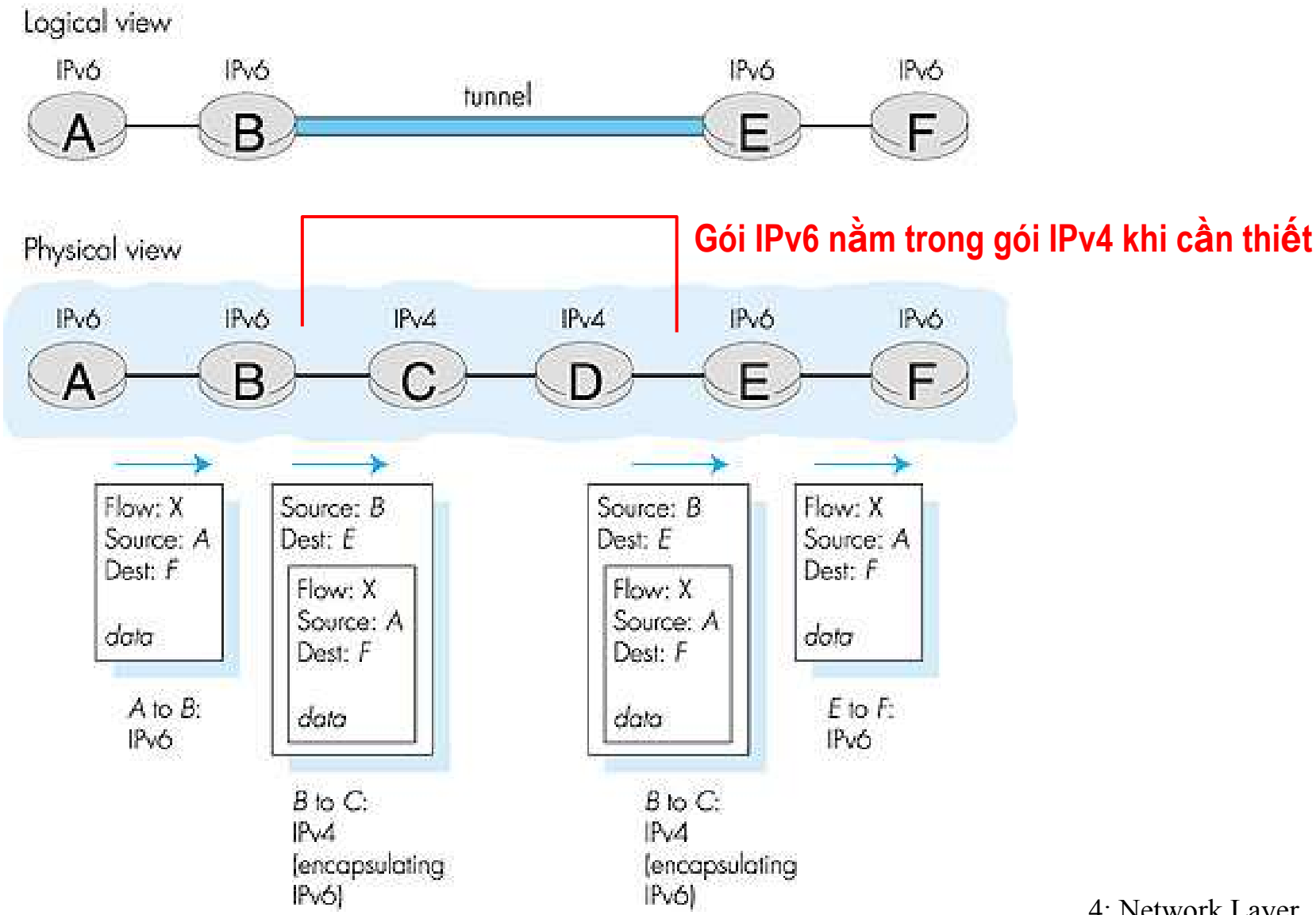
Chuyển từ IPv4 sang IPv6

- Không thể đồng thời nâng cấp tất cả router
  - Không chọn được ngày
  - Làm thế nào để tồn tại cả hai Hệ thống IPv4 và IPv6?
- Hai giải pháp được đề xuất:
  - *Dual Stack*: Một số router hỗ trợ cả 2 bộ giao thức (v6, v4) để "biên dịch" giữa hai khuôn dạng
  - *Tunneling:* Gói tin IPv6 được đặt trong trường dữ liệu của gói tin IPv4 khi truyền giữa các router IPv4

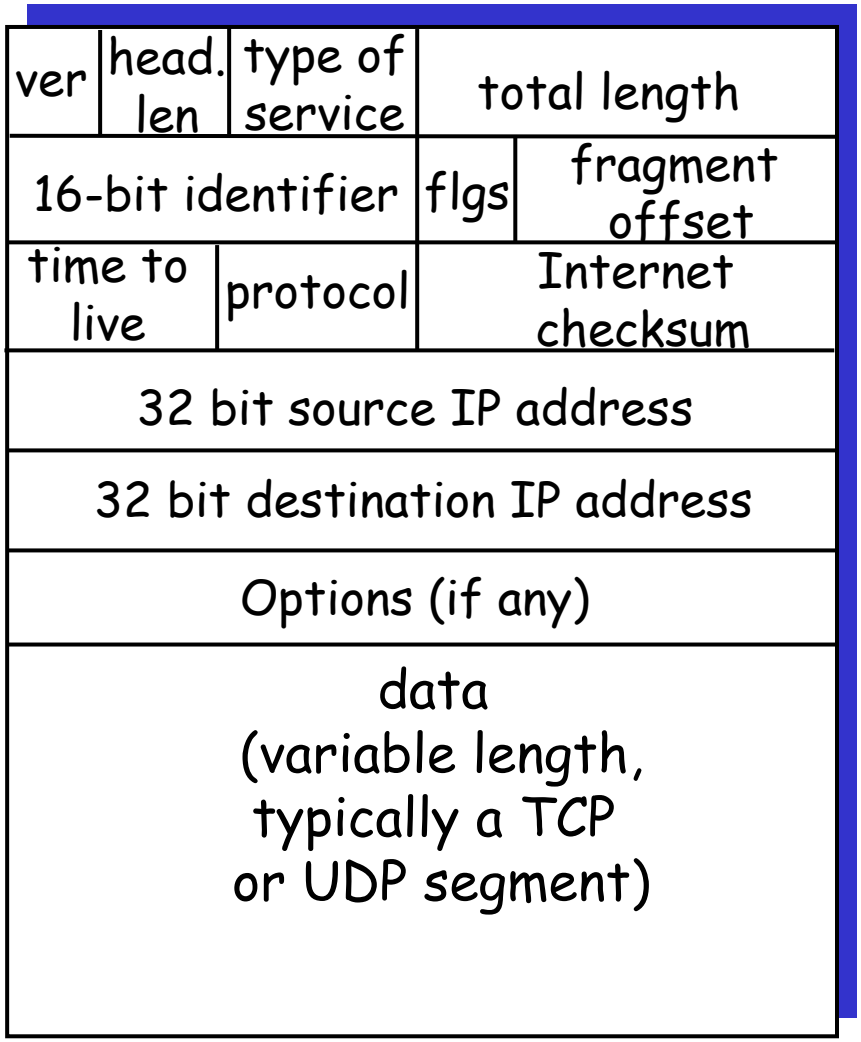
Giải pháp Dual Stack

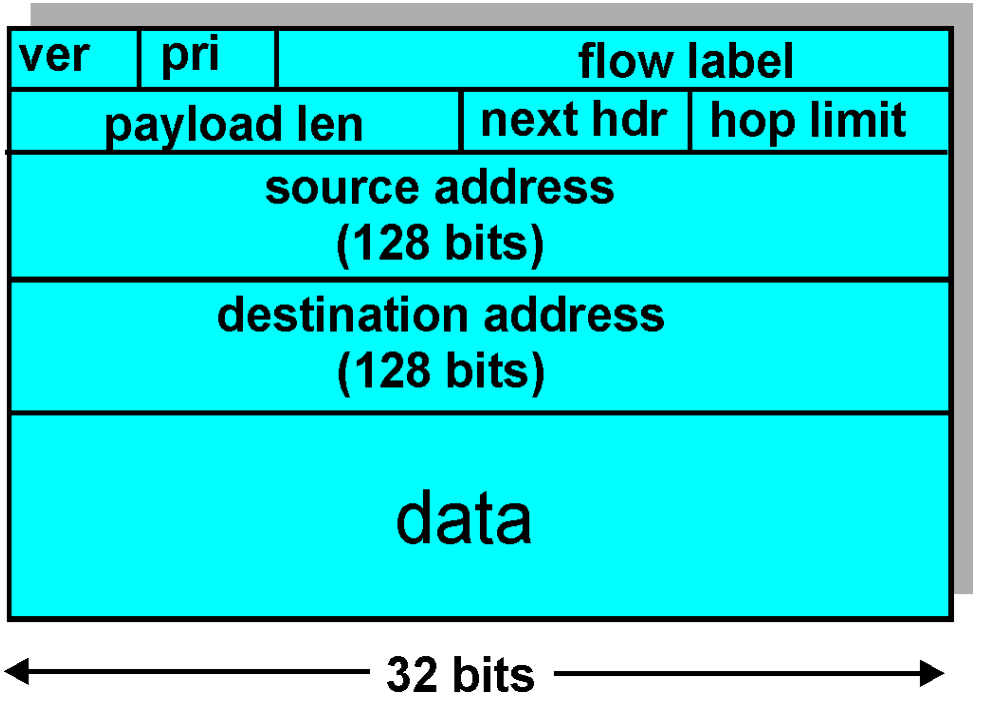


Đường ống (Tunneling)



IPv4 vs. IPv6


ver
head. len
type of service
total length
16-bit identifier
flgs
fragment offset
time to live
protocol
Internet checksum
32 bit source IP address
32 bit destination IP address
Options (if any)
data
(variable length,
typically a TCP
or UDP segment)


ver
pri
flow label
payload len
next hdr
hop limit
source address
(128 bits)
destination address
(128 bits)
data
32 bits